RUNNING HEAD: Role of logographemes in Chinese handwriting

Exploring the role of logographemes in Chinese handwritten word production

Markus F. Damian[1] and Qingqing Qu[2]

[1]School of Experimental Psychology, University of Bristol, United Kingdom

[2]Key Laboratory of Behavioral Science, Institute of Psychology,

Chinese Academy of Sciences, Beijing, China

Word count: 7,145 for main text (excl. Abstract and References)

Address for correspondence:

Qingqing Qu
Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences, Beijing, China
16 Lincui Road, Chaoyang District, Beijing, China
100101
China
Tel: +86-10-64888629
Fax: +86-10-64872010
Email: quqq@psych.ac.cn

Abstract

Recent research has demonstrated that abstract orthographic representations such as morphemes, syllables, and graphemes, influence handwritten production in languages with alphabetic scripts. The orthographic representations involved in the written production of non-alphabetic languages such as Chinese are less well understood. Chinese words consist of one or more characters which typically contain embedded radicals, with radicals themselves composed of strokes. A logographemic representational level, in between radical and strokes, has also been postulated. Here we report four experiments using a form preparation task ("implicit priming") to test for the presence of radical and logographemic priming effects in writers of simplified Chinese characters. We found strong evidence for radical-based effects, but only weak evidence for logographemic priming effects, which contrasts with recent positive logographemic priming effects reported by Chen and Cherng (2013) for writers of traditional characters. Possible reasons for this discrepancy are discussed in terms of potential differences between simplified and traditional scripts, as well as other procedural differences.

Keywords: handwriting; orthographic production; Chinese; logographemes

Exploring the role of logographemes in Chinese handwritten word production

To write a word, one first activates a word's meaning in the semantic system, subsequently retrieves its corresponding lexical codes from the orthographic output lexicon and then retrieves allographic representations and graphic motor patterns before the neuromuscular movement is performed. A fundamental question in the written production literature concerns the structure of orthographic representations. Do the orthographic codes merely consist of abstract letter identities, or apart from the identity (and serial order) of letters do they also involve higher-level linguistic representations such as morphemes, syllables, and graphemes? Neuropsychological case studies and chronometric studies conducted with unimpaired individuals indeed suggest that higher-level abstract representations constrain orthographic output processes (see e.g., Caramazza & Miceli, 1990 for a neuropsychological study; Kandel, Álvarez, & Vallée, 2008; Orliaguet & Boë, 1993 for effects of morphemes; Lambert, Kandel, Fayol, & Espéret, 2008; Kandel, Álvarez & Vallée, 2006, for syllables; Kandel & Spinelli, 2010; Kandel, Soler, Valdois, & Gros, 2006; Shen, Damian & Stadthagen-Gonzalez, 2013, for abstract graphemes).

Compared to Western European languages with alphabetical scripts, relatively less work has explored the structure of orthographic representations underlying Chinese written word production. Different from most European languages, Chinese has a logographic writing script. A Chinese character (also called "Han zi" or Han character) maps onto a spoken syllable and most of the time maps onto a morpheme. Therefore, Chinese is also described as a morphosyllabic language. In general, the orthographic system of Chinese is described by the following levels: words, characters, radicals, and strokes. For example, 汉语 (Chinese, /han4yu3/) is a disyllabic word that is composed of two characters (汉, /han4/, Chinese & 语, /yu3/, language), which in turn, are composed of one or more radicals (e.g., 语 consists of two radicals "讠" &"吾"), which, in turn, are composed of strokes. Strokes, as the smallest units of Chinese orthography, are described as a single movement during writing: from the point the pen touches the paper to the point the pen leaves the paper (Law & Leung, 2000). There are eight basic strokes which generate 29 compound strokes. Each character has a specific stroke order. Children are taught strokes and stroke order rules before being taught character reading and writing.

In a data corpus of the Chinese lexicon (2003) which includes more than 90,000 words, the majority of words are disyllabic (64%) and hence consisting of two characters, or multisyllabic (33%) consisting of more than two characters, whereas monosyllabic words only account for 3%. There are more than 20,000 characters in the modern Chinese language, including about 3,000 commonly used characters (Han, Zhang, Shu, & Bi, 2007). These characters are combined with each other and form much a larger amount of words.

Approximate 85% of Chinese characters are so-called *phono-semantic compounds* (Zhu, 1988) which are composed of a semantic radical that suggests the meaning of the character, and a phonetic radical that provides cues to its pronunciation. For example, in the character “语” (language, yu3), the semantic radical “讠”, which means “something related to speaking”, provides semantic information, and the phonetic radical “吾” (wu3) shares the same rhyme and tone with the whole character “语” (yu3). As shown in this case, the sound of the phonetic radical is not identical to the sound of the character. In fact, only 38% of phonetic radicals provide the correct pronunciation of characters while ignoring tones (e.g., the sound of the phonetic radical “马”, /ma3/, horse, is the same as “妈”, /ma1/, mother), and sometimes the phonetic radical has an entirely different sound from the character (e.g., the sound of “又”, /you4/, is different from “汉”, /han4/; Zhou, 1978). Moreover, the positions of radicals in characters are usually not arbitrary. Some radicals such as “讠” only appear on the left side in horizontal characters, and some radicals such as “吾” only appear on the right side. Chinese characters in a sentence or text are evenly spaced regardless of word boundaries.

A growing body of literature suggests that during visual word recognition, Chinese characters are automatically decomposed into subcharacter components (e.g., Ding, Peng & Taft, 2004; Taft, Zhu, & Peng, 1999; Zhou & Marslen-Wilson, 1999). In the written production literature, a few studies also indicate that Chinese writing involves independent radical processing (e.g., Law, 1994; 2004; Law & Caramazza, 1995; Law, Yeung, Wong, & Chiu, 2005; Qu, Damian, Zhang, & Zhu, 2011; Qu & Damian, 2015; Qu, Damian, & Li, 2016).

Logographemes as representational units in Chinese orthography?

However, not all Chinese characters consist of semantic and phonetic radicals and remarkably, some semantic and many phonetic radicals can be further divided into smaller components. Therefore, a number of linguists and psycholinguists (e.g., Fu, 1991; Law & Leung, 2000; Su, 1994; Zhang, 1984) have proposed a further representational unit to describe Chinese orthography, namely the *logographeme*, i.e., a unit larger than a stroke but smaller than a radical. For example, in the character "语" (/yu3/, language), the right phonetic radical "吾" can be further divided into two logographemes "五" (/wu3/, five) and "口" (/kou3/, mouth) while the left semantic radical cannot be further divided into logographemes and thus itself corresponds to an independent logographeme. The same concept has been proposed as 部件, /bu4jian4/, components, in mainland China in earlier linguistic references (CCCSGCSIP, State Language Commission, 1998). Only the smallest blocks that cannot be further decomposed into other logographemes are considered logographemes. Logographemes are distinctive if they have different meaning even they have similar appearances. For example, the two visually similar components 人 (/ren2/, people) and 入 (/ru4/, enter) are identified as two different logographemes because they have different meanings. Traditional Chinese characters are mainly used in Taiwan, Hong Kong and Macau. In a corpus of logographemes based on characters appearing in Hong Kong primary school textbooks, Lui, Leung, Law, and Fung (2010) identified 249 logographemes, 151 of which had no meaning and sound, 14 of with meaning only, and 84 which had associated meaning and sound. For the simplified Chinese characters predominant in Mainland China, the Chinese Character Component Standard of GB13000.1 Character Set for Information Processing (CCCSGCSIP, State Language Commission, 1998) lists a total of 560 logographemes. Unlike characters and radicals, logographemes are typically not explicitly taught in school.

Law and Leung (2000) proposed that the logographeme is the basic unit of Chinese writing, and suggested that the Chinese orthographic system should be characterised as a stroke-logographeme-radical-character hierarchy. Evidence from brain-impaired dysgraphic patients lends some support to the possibility that the logographeme constitutes a basic orthographic processing unit underlying Chinese written production. Law and Leung (2000) reported Cantonese patient SFT, whose writing errors mostly took place at the level of logographeme, and on this basis argued that the logographeme is the basic unit of Chinese orthographic representation. Law and Leung's argument was subsequently questioned by Han et al. (2007) because SFT was seriously impaired not only in delayed copying (40% correct) but also in direct copying (53% correct), and the patient clearly performed much better on an oral than a written naming task (75% vs. 25% correct). This pattern, according to Han et al., raises the possibility that the patient's errors might be attributable not only to high cognitive levels but also to deficits at a peripheral stage. Moreover, a large portion of the errors classified as involving logographemes were confounded with other error types, such as phonological (坡, /po1/, slope→破, /po4/, to break) and semantic (跑, /pao3/, to run → 路, /lu4/, road) substitutions. A large portion of so-called logographeme errors could also be classified as errors involving radicals (e.g., 坡, /po1/, slope→ 破, /po4/, to break).

Han et al. (2007) themselves reported a brain-damaged Mandarin speaker WLZ who showed perfect performance in a direct copying task yet was severely impaired in delayed copying, which indicated that his writing errors were not due to a peripheral motor deficit. His errors also did not originate at the lexical level, as evidenced by the fact that errors were unaffected by lexical and semantic factors. Based on these observations, the patient's deficit was assumed to reside between the orthographic lexicon and the motoric level, i.e., in the so-called *graphemic output buffer* proposed in Rapp and Caramazza's (1997) model. Because the patient's most prevalent errors fit with logographemes and errors were modulated by the frequency and number of logographemes, the authors proposed that in Chinese orthographic production, the graphemic output buffer is structured according to logographemes ("*logographeme output buffer").*

To our knowledge, the only evidence with regard to logographemes which stems from unimpaired Chinese writers rather than from patient case studies was recently reported by Chen and Cherng (2013). Because their experimental technique is also used in the experiments reported below[1], we will review the task in some detail in the following section.

The "implicit priming" task

Over the last 25 or so years, the implicit priming (IP) paradigm has been widely used to investigate spoken word production (e.g., Meyer, 1990, 1991; Roelofs, 1996, 1998, 1999; Roelofs & Meyer, 1998). The task involves repeatedly producing response words from among very small sets of words, and the central observation is that spoken responses are faster when response words within a set have word-initial phonological overlap ("homogeneous condition"), compared to when the same words are produced in sets in which there is no such overlap ("heterogeneous condition"). In its simplest form, the task involves spoken object naming (e.g., Roelofs, 1999, Experiment 3; Roelofs, 2006, Experiment 2; Alario, Perre, Castel & Ziegler, 2007; O'Seaghdha & Frazer, 2014, Experiment 1). However, using depictable entities as targets places severe constraints on stimulus selection, and therefore, more

[1]Indeed, our Experiments 1-3 were conducted before Chen and Cherng's (2013) work was published.

typically responses in IP tasks are elicited based on memorised associated word pairs (“prompts” and “responses”). Participants first learn small sets of word pairs such as *hot-cold, dog-cat, blood-clot*. During the following experimental block, on a trial, a prompt is presented (e.g., hot) and participants speak aloud the corresponding response word as fast as possible (cold). Independent of how responses are elicited, the basic finding from IP studies is that when response words share word-initial phonological elements (in homogeneous blocks), response latencies are faster than when they do not (in heterogeneous blocks).

This task has been used in a variety of studies to explore how speakers plan spoken responses. Meyer (1990, 1991) found that shared initial syllable/segments yielded a benefit in response latencies and that the benefit increased with increasing overlap, whereas shared non-initial syllable/segments did not. Results from the IP task have played a crucial role in constraining the WEAVER++ model of spoken production (Levelt, Roelofs & Meyer, 1999), and they have been the subject of extensive computational modelling (Roelofs, 1997). The model proposes a “suspend-resume” mechanism according to which in an implicit priming context, speakers plan part of a word form in advance as far as possible and then “suspend” planning, and they resume processing only once the response word has been identified. When response words share part of phonological form, as in the homogeneous condition, the shared portion can be prepared in advance before the specific response has been identified, whereas this is not possible in the heterogeneous condition, and hence a priming effect emerges in response latencies in the former compared to the latter. Because phonological encoding takes place incrementally (“from left-to-right”), priming will only be found if overlap between words occupies the word-initial position. Other aspects which have been investigated with this task are how the metrical structure of a word form is planned (e.g. Roelofs & Meyer, 1998), the respective role of segments vs features (Roelofs, 1999), the possibility of orthographic effects in spoken word production (e.g., Damian & Bowers, 2003; Alario et al., 2007), and many others.

Implicit priming effects have also been recently documented in written, rather than spoken, word production. Here, rather than measuring the acoustic onset of a spoken word, participants write response words on a digital graphic tablet, and the time of first contact of the stylus with the writing surface provides the response latency. For instance, Afonso and Álvarez (2011) used the IP task with Spanish writers to explore whether written production is influenced by phonological properties of response words. Shen, Damian and Stadthagen-Gonzalez (2013) provided some evidence for the claim that writers plan abstract graphemes when preparing written responses.

In a recent study, Chen and Cherng (2013) used the IP task in association with written responses to explore relevant planning units of Chinese written production. In a series of experiments, response words were selected such that critical sets of words in the homogenous conditions overlapped in terms of units of different sizes. They found no implicit priming when overlap was restricted to the first, or the first two, strokes. By contrast, priming was obtained when the first radical overlapped among response words, and most importantly, priming was also found when logographemes overlapped among responses. The authors introduced a model of Chinese handwritten character generation which adopts a similar architecture to the spoken production model described in Levelt et al. (1999). Their model focuses on the character as the main unit of written production, even though the majority of Chinese words consist of more than single character. It consists of two separate and serial stages of form encoding, “morphological encoding” and “orthographic encoding”. Morphological encoding consists of the decomposition of a complex character into radical and non-radical components, with components assigned to slots in a morphological frame. During orthographic encoding, logographemes are specified and associated to a structural frame according to the orthotactic principles of written Chinese, with the results consisting of a square-shaped "orthographic character" which is then subsequently allologographemically and motorically encoded for production. Hence, in this model the logographeme represents the "proximate unit" of handwriting, i.e., the first and most salient form-related selectable unit below the lexical level (O'Seaghdha, Chen & Chen, 2010).

The current study

In the experiments reported below, we aimed at further exploring the potential role of logographemes in Chinese written word production. There are at minimum two aspects which deserve investigation: first, participants in Chen and Cherng's (2013) study, Taiwanese students, used the so-called "traditional" Chinese script, with a set of characters which has been in use in more or less unchanged form since the 5th century. Traditional characters are in use in Taiwan, Hong Kong, and Macau. However, mainland Chinese individuals as well as citizens of Singapore and Malaysia have been using a "simplified" character set since the 1950s and 1960s in which the number of strokes has been substantially decreased, and complexity is reduced. Traditional and simplified corresponding characters are often substantially different from each other. At minimum, we should therefore investigate whether the logographeme-based priming documented by Chen and Cherng is also present when individuals write words in simplified characters (several native writers of traditional or simplified script offered their intuition to us that logographemes could be more salient in the traditional characters).

Second, Chen and Cherng (2013) used the implicit priming task in a somewhat unusual fashion. As summarised in the previous section, in the spoken domain the canonical form of this task involves participants memorising associated word pairs, with the first member ("prompt") used to elicit the second ("response"). This is also the form in which the task was used in previous investigations of written, rather than spoken, word production (Alonso & Álvarez, 2011; Damian & Stadthagen-Gonzalez, 2009; Shen, Damian & Stadthagen-Gonzalez, 2013). By contrast, Chen and Cherng chose two-character response words as targets, where the first character served as the cue, and the response consisted of writing the second character. It is not impossible that this form of the task might have promoted a visualisation of the target orthography which may have promoted the found logographeme effects, but which may not be typical of the implicit priming task in its more canonical form (nor perhaps of writing outside the laboratory).

In a first experiment, and before tackling the critical issue of the role of logographemes, we attempted to capture the effects of word-initial *radical* overlap in writers of simplified Chinese characters. To reiterate, radicals clearly represent important functional units of Chinese orthography, and Chen and Cherng (2013) reported radical-based priming effects of 19 and 57 ms in two experiments (the authors speculated that the different in size between the two studies may have arisen from differences in materials in terms of stroke complexity). We used a form of the IP task which closely resembled the one used in numerous previous studies on spoken and written production.

Experiment 1

Method

Participants. Twenty students from Beijing Forest University and China Agricultural University participated in the experiment and were paid approximately £5. All were native Mandarin Chinese speakers and writers of simplified Chinese characters. All had normal or corrected-to normal vision.

Materials and Design. The independent variable “Context” included two conditions: homogeneous and heterogeneous. In the homogeneous condition, the left characters of disyllabic response words within an experimental block shared the initial radical. In the heterogeneous condition, response words within an experimental block did not share orthographic form.

For the *homogeneous* condition, four sets of four pairs of disyllabic response words were selected so that the first characters of disyllabic response words shared the initial radical. All response words were semantically and phonologically unrelated to each other in the present study. Four *heterogeneous* sets with four disyllabic response words in each set were then formed by swapping items between the homogeneous sets so that the response words did not share the initial radical (see Appendix A for a complete list of stimuli). In this way, each item was presented exactly once in each set, hence, the same stimuli were used in the two conditions, and the context was the only difference. Response words had a mean frequency of 11.5 per million in the database of Chinese Lexicon (2003) and consisted on average of 20 strokes. For each response word, a corresponding highly associated disyllabic prompt word was selected to form a word pair (e.g., 工作, /gong1zuo4/, job - 聘用, /pin4yong4/, hire).

Each participant carried out a total of eight testing blocks (four homogeneous blocks and four heterogeneous blocks), following a practice block. Within each block, each of the four prompt words was repeated five times, generating a total of 20 trials. The entire experiment included 160 testing trials. The order of the blocks was randomised. The items within each block were pseudo-randomised using the software Mix (van Casteren & Davis, 2006), with the constraint that a prompt word was never repeated on adjacent trials. A new order of random trials and blocks was generated for each participant.

Apparatus. Stimuli were presented from an IBM-compatible computer on a 17 inch monitor using DMDX 3.0 (Forster & Forster, 2003). Prompt words were presented in Arial bold font of size 18. The stimuli were displayed toward the bottom of the screen to minimise head movements between displays and the writing surface. Written responses were collected using a WACOM Intuos3 A4 graphic tablet, an Intuos ink digitiser pen and a double-sided A4 sheet attached to the tablet.

Procedure. Participants were first instructed to lift the pen very slightly from the answering sheet so that response could be given as fast as possible; they should not drop the pen on the sheet before a response had been identified. They were encouraged to write as fast and accurately as they could, but not to correct their answer if a mistake was made. If a word was forgotten, they were told that the line should be left empty. Participants were then informed that the experiment consisted of a series of relatively short experimental blocks. Their task was to memorise four associated word pairs prior to the beginning of each block; during the following block, they should write down response words as fast and correctly as possible on the answer sheet when cued by the prompt word. Participants could spend as long as necessary to memorise the word pairs for each block.

The session began with a practice block of 12 trials in which items other than those included in the experiment were used. Subsequently, the eight experimental blocks were administered. Participants were allowed short breaks between blocks. On each trial, participants saw a sequence consisting of a fixation cross (500 ms), a blank screen (500 ms), a prompt word (4,000 ms), and an inter-trial interval (1500 ms). The entire experiment took approximately 30 minutes per participant.

Results

Response sheets were inspected for errors, and response latencies were extracted for each trial. Trials with errors (2.3%) and trials with response latencies longer than 1,800 ms or shorter than 200 ms (2.4%) were eliminated from the latency analysis (total: 4.7%). Table 1 shows the mean response times and error rates of each condition. As can be seen, radical overlap generated a priming effect of 69 ms on latencies.

The results were analysed using a linear mixed effects approach (Baayen, Davidson & Bates, 2008; Bates, 2005) which simultaneously takes participants and items variability into account, using the software R (R Development Core Team, 2015) with the package lme4 (Bates & Maechler, 2016). Model fitting was carried out by initially specifying a model that only included adjustments to the intercept for participants and items as random effects. The initial model was enriched by adding the fixed factor Context. We additionally fitted models with individual by-participant and by-item slopes for the fixed effect Context.[2] The results showed that the best-fitting model included slope adjustments to both participants and items, $\chi 2(4, N = 3{,}429) = 85.28$, $p < .001$, compared to a model without such slope adjustments. In this best-fitting model, the fixed factor Context was highly significant, $\beta = 72$, $SE = 23.8$, $t = 3.01$, $p = .006$.

Parallel analyses were conducted on the errors, but a mixed logit model was used due to the categorical nature of the data (Jaeger, 2008). The only model which converged contained Context as a fixed factor, and by-participant intercept and slope adjustments, but only by-item intercept adjustments. In this model, Context was not significant, *Wald z* = 1.08, $p = .279$.

Discussion

This experiment rendered the expected priming effect based on radical overlap among written response words. Latencies were 69 ms faster when writers were able to carry out partial planning based on radicals, an effect which compares well with the one reported by Chen & Cherng (2013). The results demonstrate that our procedure is able to capture priming effects in handwritten production, and more broadly they underscore the importance of radicals as central planning units of Chinese written production. In the next experiment, we attempted to capture a priming effect based on logographeme overlap among written response words.

[2]We are aware of the argument that the model structure in linear mixed effect models should always be "maximal" (Barr, Levy, Scheepers & Tily, 2013), as well as the more recent counter-argument that parsimonious models are often preferable (Bates, Kliegl, Vasishth & Baayen, 2015). In our case, as will be shown there is specific information derived from identifying the best-fitting model, so here we took the "model building" approach.

Experiment 2

Method

Participants. Twenty-four students from the University of Bristol participated in the experiment. All were native Mandarin Chinese speakers and writers of simplified Chinese characters. All had normal or corrected-to normal vision. They were rewarded £5 for their participation.

Materials and Design. As in the first experiment, the independent variable “Context” included two conditions: homogeneous and heterogeneous. In the homogeneous condition, the left characters of disyllabic response words within an experimental block shared the initial logographeme. Four logographemes (口, 日, 十, 人) were used, one for each set. It should be noted that although these logographemes embedded in response words were independent Chinese characters with independent meaning and sound (e.g., “口”, /kou3/, mouth), all response words were nevertheless semantically and phonologically unrelated to each other in the present study (e.g., 别墅, /bie2shu4/, villa - 勋章, /xun1zhang1/, medal). In the heterogeneous condition, sets with four disyllabic response words in each set were then formed by swapping items between the homogeneous sets so that the response words did not share the initial logographemes (see Appendix B for a complete list of stimuli). Response words had a mean frequency of 8 per million in the database of Chinese Lexicon (2003) and consisted on average of 18 strokes. For each response word, a corresponding highly associated disyllabic prompt word was selected to form a word pair.

As in the first experiment, each participant carried out a total of eight testing blocks, following a practice block. Within each block, each of the four prompt words was repeated five times, hence the entire experiment included 160 testing trials. The order of the blocks, and the order of trials within each block, were randomised.

Apparatus and Procedure. These were the same as in the first experiment. The session began with a practice block of 12 trials in which items other than those included in the experiment were used. Subsequently, the eight experimental blocks were administered. The entire experiment took approximately 30 minutes per participant.

Results

Trials with errors (4.2%), and with response latencies longer than 1,800 ms or shorter than 200 ms (2.6%), were eliminated from the latency analysis (total: 6.8%). Table 1 shows the mean response times and error rates of each condition. Word-initial logographeme overlap generated a priming effect of 21 ms on latencies.

Linear mixed effects analysis of the latency results showed that compared to a baseline model which included only by-participant and by-item adjustments to the intercept, a model which additionally included the fixed effect Context (Homogeneous vs. Heterogeneous) was vastly preferable, $\chi 2(1, N = 3{,}429) = 8.19$, $p = .004$. However, by adding by-participant and by-item adjustments to the slope of Context, we found that the best-fitting model included adjustments to both participants and items, $\chi 2(4, N = 3{,}429) = 85.28$, $p < .001$, when compared to the model without such slope adjustments. Indeed, in this best-fitting model, the fixed factor Context was no longer significant, $t = 1.14$, $p = .266$ (p values were calculated using the package lmerTest; Kuznetsova, Brockhoff & Christensen, 2016). Bayesian analysis showed a Bayes Factor BF_{01} of 3.45, offering "moderate" support for the null hypothesis that Context did not affect latencies (Jeffreys, 1961). The fact that inclusion of random slopes improved the fit points to considerable variation among participants and items regarding the effects of Context.

To explore this pattern, we generated plots (adopted from Baayen, 2008, p. 250) which display response latencies for individual participants and items dependent on Context. Figure 1 (top panel) shows such a plot. Subpanels correspond to participants (1-24). Dashed lines represent the model which assigns the same slope to all participants; solid lines represent the model which allows slopes to vary among participants. It is clear that the effect varies considerably between participants, with some (e.g., 3 and 5) showing a strong logographeme priming effect, but with others (e.g., 20 and 22) showing a strong reverse effect, and many (e.g., 6 and 19) exhibiting only minor or no effect. Figure 1 (bottom panel) shows a similar plot, but now for the items (1-16). Here, the solid line within each subpanel reflects the model which allows slopes to vary among items, and it is clear that when compared to the by-participant figure, the effect is relatively evenly distributed across items, with the item slopes (solid lines) quite similar to the overall slope (dashed line).

The number of previous studies which had used the IP task in association with written responses is still limited, and response time levels appear quite variable across these studies (e.g., a heterogeneous mean of 666 ms in Experiment 1 of Damian & Stadthagen-Gonzalez, 2009, a mean of 1268 ms in Experiment 1 of Afonso & Álvarez, 2011; a mean of 733 ms in Experiment 1 of Shen, Damian & Stadthagen-Gonzalez, 2013). The heterogeneous mean of Experiment 1 reported above (789 ms) is notably slower than those in Chen and Cherng (2013) who reported mean heterogeneous latencies of 667 ms when averaged across all five experiments. Perhaps the difficulty to find significant logographemic priming in our first experiment is related to the slower responses, and the inconsistency in the results among participants (see Figure 1) might be explained in terms of differences in response speed. We explored the possibility that the size of the logographemic overlap effect is related to overall speed in the task by generating a so-called caterpillar plot adopted from work by Kliegl and colleagues (e.g., Kliegl, Masson & Richter, 2010). Figure 2 shows on the left panel the intercept adjustment for each participant, sorted by overall speed. Participants with a large negative adjustment for the intercept are fast responders, whereas participants with a large positive adjustment are slow responders; hence the plot shows the fastest participant at the bottom, and the slowest on top. The right panel shows the corresponding priming effect as reflected in by-participant slope adjustments; intercept and slope adjustments were obtained by using the function ranef(). There is no obvious relation between speed and size (or direction) of the priming effect in our data set.

Parallel analyses were conducted on the errors, but a mixed logit model was used due to the categorical nature of the data (Jaeger, 2008). Compared to a baseline model which included only by-participant and by-item adjustments to the intercept, a model which additionally included Context provided a better fit, $\chi 2(1, N = 3{,}677) = 5.72$, $p = .017$. As for response latencies, the best-fitting model included adjustments to the slopes of Context both for participants and by items, $\chi 2(4, N = 3{,}677) = 29.67$, $p < .001$, and in this model, the fixed factor Context was no longer significant, *Wald z* = 0.514, $p = .607$.

Discussion

Response latencies were faster when response words shared an initial logographeme relative to when they did not. This facilitation effect lends some support to the logographeme as an orthographic representational unit. Yet, the effect was numerically small (21 ms, compared to the 50 and 34 ms logographeme priming effects reported by Chen & Cherng, 2013) and not particularly stable across participants and items. Hence, the important finding from Experiment 2 is that effects of logographeme are more difficult to obtain (with our participants, target script, and IP method) than one might have predicted based on Chen and Cherng's results.

More errors were observed in the homogeneous than in the heterogeneous condition, a pattern which is consistent with previous findings from the IP task in the spoken production literature (e.g., Chen & Chen, 2007; Chen, Chen, & Dell, 2002; Meyer, 1991). However, as was the case for response latencies, the error effect was not statistically reliable.

In the next experiment, we made another attempt to obtain logographeme priming with independent materials, but we additionally tested the possibility that implicit priming with written Chinese response words could result purely from motoric redundancy: perhaps in homogeneous blocks, the fact that responses share word-initial strokes and hence corresponding motor codes might be enough to generate a response time benefit. Note that Chen and Cherng (2013) also considered this possibility, by conducting experiments in which target characters overlapped in one or two word-initial strokes, and found no priming effect. But given the difference in target scripts and method, we decided to put this possibility to the test in Experiment 3.

Experiment 3

In Experiment 3, in addition to the homogeneous and the heterogeneous condition, an *inconsistent* condition was included in which response words shared visually and motorically similar orthographic constituents but these actually corresponded to different logographemes (e.g., 亻 vs. 彳). If the weak implicit priming effect found in Experiment 2 arises entirely at the motoric level, then motorically similar orthographic constituents should be sufficient to cause a facilitation effect, namely, the inconsistent condition should be significantly faster than the heterogeneous condition. By contrast, if the implicit priming effect is not attributable to motoric redundancy but is rather due to the planning of abstract logographemes, no priming effect should arise in the inconsistent condition.

Method

Participants. Twenty students from the same population as Experiment 1 participated in the experiment. All were native Mandarin Chinese speakers and writers of simplified Chinese characters, and all had normal or corrected-to normal vision. They were rewarded £5 for their participation.

Materials and Design. The target stimuli consisted of 12 disyllabic response words in four sets with three members in each set. Four logographemes (日 , 口, 士, 十) were selected as the overlapping part. These four logographemes formed two pairs [日 vs. 口] and [士 vs. 十] and the two logographemes in each pair were visually similar and shared most strokes. In the *homogeneous* sets, the first character of disyllabic responses words within each of four sets shared an initial logographeme. The 12 response words had a mean frequency of 9.83 per million in the database of Chinese Lexicon (2003) and an average stroke number of 18. *Heterogeneous* sets were generated by swapping one item between the homogeneous sets, such that the first characters of response words no longer shared the initial logographemes. *Inconsistent* sets were generated by again exchanging one item with another one from a different set, so that the first characters of response words were visually similar, but different logographemes occupied initial position. See Appendix C for a complete list of materials. For each response word, a corresponding highly associated prompt word was selected. All prompt and response words were nouns. Semantic or associative relations between prompts or responses were avoided.

Each participant carried out a total of 12 testing blocks, following a practice block. Within each block, each of the four items was presented five times, for a total of 20 trials. The order of the blocks was randomised for each participant, and within each block the sequence was pseudo-randomised such that a prompt word was never repeated on adjacent trials. The entire experiment comprised 240 testing trials for each participant and took about 40 min per participant to complete.

Apparatus and Procedure. These were the same as in the first two experiments.

Results and Discussion

Trials with errors (2.3%), and with response latencies longer than 1,800 ms or shorter than 200 ms (2.7%), were eliminated from the latency analysis (total: 5.0%). Table 1 shows the mean response times and error rates of each condition. Response latencies appear very similar across the three experimental conditions.

Latencies were analysed using a linear mixed effect model. Compared to a baseline model which included only by-subject and by-item adjustments to the intercept, a model which additionally included the fixed effect of Context did not significantly improve the fit, $\chi 2(2, N = 4{,}561) = 2.40$, $p = 0.302$. In the latter model, Context corresponded to a BF_{01} of 4.29.

Errors, analysed with a mixed logit model, showed that compared to a baseline model with by-subject and by-item adjustments to the intercept, a model with the fixed effect of Context did not render a significant improvement in fit, $\chi 2(2, N = 4{,}800) = 0.076$, $p = .963$.

Hence, the results obtained in Experiment 2 offer a resounding null finding, with absolutely no trace of a logographeme-based implicit priming effect. The discrepancy between our weak or null findings and those previously reported by Chen and Cherng (2013; to reiterate, these authors found priming effects of 50 and 34 ms in two experiments) is certainly baffling, and hence it is worth highlighting differences between the two sets of studies which might have contributed.

The first and foremost, already mentioned in the Introduction, is that participants in Chen and Cherng's (2013) experiments wrote traditional Chinese characters, whereas in our studies, they wrote simplified characters. It is not impossible that systematic differences in whether or not logographeme-based priming effects are obtained have to do with the nature of the response script. A further substantial difference lies in the method to elicit responses (again, see Introduction): in the experiments reported above, we used the "canonical" form of implicit priming in which "prompt" words are used to elicit production of associated "response" words, whereas Chen and Cherng presented the first character of two-character words to elicit written production of the second character. A third - and perhaps minor - difference is that in the experiments reported above, blocks corresponding to the experimental conditions were presented in a different random order to each participant. By contrast, Chen and Cherng arranged experimental blocks such that homogeneous and heterogeneous blocks appeared in separate halves of the experiment, with the order of the two halves counterbalanced across participants. Perhaps, the fact that in this form of the task, homogeneous blocks appear one after the other highlights the salience of the experimental manipulation more so than the random block order chosen in our experiments.

In the fourth experiment, we addressed issues two and three above: we adopted the 1^{st}-to-2^{nd} character elicitation method used by Chen and Cherng (2013), and we blocked experimental conditions in separate halves of each testing session, counterbalanced across participants.

Experiment 4

Method

Participants. Twenty-four students from Beijing Forest University and China Agricultural University participated in the experiment and were paid approximately £5. All were native Mandarin Chinese speakers and writers of simplified Chinese characters. All had normal or corrected-to normal vision.

Materials and Design. 16 disyllabic target words were chosen such that four of them shared the initial logographeme of the second character (e.g., 离别, 功勋, 卑鄙, 子嗣), hence forming four homogeneous sets of four items. Four heterogeneous sets were formed by combining the target words such that the initial logographeme of the second character did not overlap. See Appendix D for a list of materials.

Each participant carried out a total of eight testing blocks (four homogeneous blocks and four heterogeneous blocks), following a practice block. Within each block, each of the four prompt words was repeated five times, generating a total of 20 trials. The entire experiment included 160 testing trials. For half of the participants, blocks were homogeneous in the first half of the experiment, and heterogeneous in the second, and vice versa for the other half of the participants. Within each half, the order of blocks was determined by a Latin Square design. The items within each block were pseudo-randomised such that a prompt word was never repeated on adjacent trials. The entire experiment comprised 160 testing trials for each participant and took about 30 min per participant to complete.

Apparatus and Procedure. These were the same as in Experiment 1 and 2, except for the elicitation method: on each trial, the first character of each response word was presented on the computer screen, and participants were instructed to write down the corresponding second character as quickly and accurately as possible.

Results and Discussion

Trials with errors (5.3%) and trials with response latencies longer than 1,800 ms or shorter than 200 ms (1.4%) were eliminated from the latency analysis (total: 6.7%). Table 1 shows the mean response times and error rates of each condition. As can be seen, word-initial logographeme overlap in the homogeneous condition induced a 21 ms implicit priming effect.

Latencies were again analysed using a linear mixed effect model. The results showed that compared to a baseline model which included only by-participant and by-item adjustments to the intercept, a model which additionally included the fixed effect of Context was vastly preferable, $\chi2(1, N = 3,582) = 11.37$, $p < .001$. Adding by-participant adjustments to the slopes further improved the fit, $\chi2(2, N = 3,582) = 241.50$, $p < .001$, but adding by-item adjustments did not, $\chi2(1, N = 3,429) = 3.42$, $p = .180$. In the best-fitting model, the fixed factor Context was no longer significant, $t = 0.98$, $p = .336$, $BF_{01} = 4.04$. As was the case in the first experiment, this pattern highlights considerable variation between participants regarding the effects of Context, which is also reflected in Figure 3 (top panel). A caterpillar plot presented in Figure 4 suggests, similar to Experiment 2, that this participant variability is not related to overall speed.

Errors, analysed with a mixed logit model, showed that compared to a baseline model with by-participant and by-item adjustments to the intercept, a model with the fixed effect of Context did not render a significant improvement in fit, $\chi2(1, N = 3{,}840) = 0.005$, $p = .941$.

In summary, the logographeme-based priming effect obtained in latencies was suggestive, but once again weak and statistically not reliable. This suggests that the two aspects which were modified in Experiment 4 relative to the earlier studies (response elicitation method, and arrangement of experimental conditions into blocks) were of minor relevance, and don't explain the discrepancy between the robust logographeme effects reported by Chen and Cherng (2013), and our repeated difficulties in obtaining the effects which would be predicted from their results.

General Discussion

The experiments reported here sought to investigate the structure of orthographic representations underlying Chinese written word production, with a focus on the possibility that logographemes (representations in between strokes and radicals) might occupy an important position. A first experiment using the implicit priming task showed that word-initial radical overlap generated a substantial priming effect. Across the following three experiments which specifically targeted logographeme- (rather than radical-) based priming, the accumulated evidence was largely negative. Although in two of the experiments, logographemic overlap among written response words resulted in a numerical effect in the predicted direction, this effect was generally statistically weak, and characterised by strong variability among participants. Across the three experiments, calculation of Bayes Factors offered moderate but consistent support for a null finding regarding logographemes.

Our difficulties in detecting logographemically based priming effects are in stark contrast to recent results reported by Chen and Cherng (2013) who in two experiments found strong and reliable effects of logographeme overlap, and took their findings to imply that logographemes occupy a central role (“proximate unit”) for Chinese writers. The most obvious difference between theirs and our experiments was that they used an elicitation technique in which first characters of two-character response words were presented and participants wrote down the corresponding second character. By contrast, in our own Experiments 1-3, we used the more standard form of the IP task in which “prompt” words were used to elicit highly associated “response” words. Perhaps this difference in methodology accounts for the discrepancy between the two sets of findings. However, Experiment 4 tested, and ultimately rejected, this possibility: even with a 1^{st}-to-2^{nd} character elicitation version of the IP task, we failed to obtain statistically reliable logographeme priming effects.

There was a remaining procedural difference which deserves to be discussed. In both sets of experiments, prompts were presented on a computer screen and participants wrote down their responses on a digital graphic tablet. In Chen and Cherng’s (2013) procedure, participants wrote with a stylus onto the tablet and their written output appeared on the computer screen below the prompt. By contrast, in our experiments, a sheet of paper was attached to the graphic tablet, and participants wrote on the paper with an inking digitiser pen; this or a similar setup is used in a growing number of articles on handwritten production (e.g., Bonin, Peereman & Fayol, 2001; Kandel et al., 2009; Shen et al., 2013). Both setups provide visual feedback about written output to participants, and it is difficult to see (although not impossible) how the difference between them could determine whether or not logographeme effects emerge.

A more likely candidate, briefly mentioned in the Introduction, is that participants in Chen and Cherng’s (2013) study wrote traditional Chinese characters, whereas our participants wrote in simplified script. In other words, the target script itself may affect the processing units underlying writing. Compared to traditional Chinese characters, simplified characters have fewer strokes. For example, the commonly used character 变 (bian4, “change”) has eight strokes in simplified form, while its traditional form “變” has 23 strokes. Moreover, some traditional characters are quite similar in appearance, such as 書 (shū) “book”, 晝 (zhòu) “daytime” and 畫 (huà) “drawing”; the simplified equivalents are 书, 昼, and 画, which look much more distinct. Perhaps due to the difference in complexity in terms of stroke numbers and similarity of appearance, writers of traditional script may utilise more refined details to represent Chinese characters. Hence it is possible – although admittedly speculative at present - that writers of traditional script are more highly sensitised to logographemes as planning units of writing than are writers of simplified script who tend to use larger units (i.e., radicals) as the primary planning units. Further research is needed to explore this potential difference dependent on script. Another possible explanation might arise from the possibility that the number of logographemes in both types of writing scripts affected the results. As outlined in the Introduction, there are an estimated 560 logographemes in the simplified script but only 249 logographemes in the traditional script, and perhaps the more numerous logographemes in simplified script render them to be less salient as proximate units of handwriting.

In addition, individual differences on learning to read and write might affect processing units in writing. Ho, Yau and Au (2003) point out that Chinese children’s literacy development differs from alphabetic language learners. One important difference is that more orthographic units need to be learned in Chinese compared to alphabetic languages. Ho et al. presented a model of orthographic knowledge development of Chinese children’s writing and reading. According to the model, Chinese children first develop knowledge about character configurations, and at this level, children can distinguish character printing from drawing. Subsequently, Chinese children are taught structural knowledge of characters, and hence acquire information about semantic and phonetic radicals, before advancing to an amalgamation stage where children acquire the skill to integrate all of these different types of orthographic knowledge about characters to write efficiently. Children are often explicitly taught the meanings of semantic radicals in school, and this improves the development of radicals as representational units. By contrast, literacy instruction on logographemes is not as uniform as with radicals. Most children are not explicitly taught logographemes whereas a minority of children are given detailed logographeme-based information. The large individual differences in learning history could cause individual differences in the degree to which writers demonstrate sensitivity to logographemes, as we found in the present study (see Figure 1 and 3). In future research, this possibility should be explicitly assessed, via self-reported information regarding learning history, as well as with more explicit assessments of individual orthographic awareness of logographemic structures.

In Figures 1 and 3, the considerable by-participants variability manifests itself as a mixture of priming (e.g., Participant 3 in Figure 3), null (e.g., Participant 2 in Figure 3), and inhibitory effects (e.g., Participant 20). If for at least some participants, word-initial logographeme overlap indeed results in an genuine inhibitory effect (rather than a null effect plus some noise), this would certainly ask for an explanation. The literature on written versions of the implicit priming task is still limited, but to our knowledge no inhibitory effects have been reported so far (e.g., Chen & Cherng, 2013, reported a mix of null and positive effects). Zhang and Damian (2012) reported a spoken implicit priming experiment with Mandarin speakers in which visual orthographic similarity among spoken response words resulted in a weak (11 ms) inhibitory effect; however, subsequent control experiments led the authors to attribute this effect to the memorisation stage of the response generation task, rather than reflecting processes genuine to speaking. In short, it is at present unclear whether the inhibitory effect of logographeme overlap which manifests itself in a subset of participants is genuine, and if so, how it could be explained.

In the Introduction, we reviewed the evidence from neuropsychological case studies with regard to the potential role of logographemes, and in particular patient WLZ (Han et al., 2007) who showed errors in a delayed copying task which mainly consisted of logographeme substitutions and deletions.

This constitutes *prima facie* strong evidence for the mental reality of the logographeme as a functional unit in Chinese writing. However, if (as our own experiments suggest), considerable individual variability exists among Chinese writers, then the results of a single case study might be less instructive than they normally are. Perhaps a future group study on Chinese dysgraphic patients will offer more comprehensive insight into whether logographemes constitute “proximate units” for Chinese spellers, or only for some of them.

To investigate the issue of logographemes in further studies, one could potentially adopt other paradigms and manipulations. One possible way is to use the picture-word interference (PWI) task. In the PWI task, form overlap between object names and distractor words generally results in facilitation. One could manipulate overlap between pictures names and distractor words in terms of logographemes (e.g., picture name: 桔子, /ju2zi/, orange; distractor word: 治理, /zhi4li3/, administer). If the logographeme is a unit of orthographic representations underlying Chinese written production, the prediction would be that distractor words which share a logographeme with the target picture names facilitate written responses. Alternatively, one could investigate the impact of logographeme frequency, or the number of logographemes, on written picture naming latencies or writing-to-dictation tasks. A difficulty is that this would require matching other lexical properties which themselves might affect written response latencies, such as written and spoken word frequency, stroke number, etc., but the number of logographemes is highly confounded with these factors. For instance, a character with more logographemes tends to include more strokes, and might also tend to have lower word frequency, than a character with fewer ones.

Overall, in the experiments reported here, we found relatively little evidence for the claim that logographemes are particularly relevant for writers of (simplified) Chinese characters. Future research will have to determine whether the discrepancy between our findings and previously reported results obtained with writers of traditional Chinese characters can be attributed to the differences between the two script systems, and why exactly logographemes should be relevant in one but not in the other.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC), No. 31400967, and the German Research Foundation (DFG) and the National Natural Science Foundation of China (NSFC) in joint project Crossmodal Learning, DFG TRR-169/NSFC No. 61621136008 to the corresponding author.

References

Afonso, O., & Alvarez, C. J. (2011). Phonological effects in handwriting production: evidence from the implicit priming paradigm*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1474.

Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102*(3), 464-475.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, 27-30.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv preprint arXiv:1506.04967.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2016). *lmerTest: Tests in linear mixed effects models.* R package version 2.0-30. http://CRAN.R-project.org/package=lmerTest.

Bates, D. M., & Maechler, M. (2016). *lme4: Linear mixed-effects models using 'Eigen' and S4.* R package version 1.1-12. http://CRAN.R-project.org/package=lme4.

Bonin, P., Peereman, R., & Fayol, M. (2001). Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language, 45,* 688–720.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition, 37,* 243–297.

Chen, J. Y., & Cherng, R. J. (2013). The proximate unit in Chinese handwritten character production. *Frontiers in Psychology, 4*(517), 10-3389.

Chen, J.-Y., Chen, T.-M. (2007). Form encoding in Chinese word production does not involve morphemes. *Language and Cognitive Processes. 22,* 1001–1020.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory & Language, 46,* 751–781.

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*(1), 119–132.

Damian, M. F., & Stadthagen-Gonzalez, H. (2009). Advance planning of form properties in the written production of single and multiple words. *Language and Cognitive Processes, 24,* 555–579.

Ding, G., Peng, D.-L., & Taft, M. (2004). The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 530–539.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers, 35,* 116–124.

Fu, Y. H. (1991). Hanzi de bujian [Logographemes of Chinese characters]. *Yuwenjianshe, 12,* 3–6.

Han, Z., Zhang, Y., Shu, H., & Bi, Y. (2007). The orthographic buffer in writing Chinese characters: Evidence from a dysgraphic patient. *Cognitive Neuropsychology, 24(4),* 341–350.

Ho, C. S.-H., Yau, P. W.-Y., & Au, A. (2003). Development of orthographic knowledge and its relationship with reading and spelling among Chinese kindergarten and primary school children. In C. McBride-

Chang, & H.-C. Chen (Eds.), *Chinese children's reading development* (pp. 51-72). Westport, CT: Praeger Publishers.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford: Clarendon Press.

Kandel, S., & Spinelli, E. (2010). Processing complex graphemes in handwriting production. *Memory & Cognition, 38(6),* 762-770.

Kandel, S., Álvarez, C. J., & Vallée, N. (2006). Syllables as processing units in handwriting production. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 18-31.

Kandel, S., Álvarez, C., & Vallée, N. (2008). Morphemes also serve as processing units in handwriting production. In M. Baciu (Ed.), *Neuropsychology and cognition of language - Behavioural, neuropsychological and neuroimaging studies of spoken and written language* (pp. 87-100). Kerala, India: Research Signpost.

Kandel, S., Hérault, L., Grosjacques, G., Lambert, E., & Fayol, M. (2009). Orthographic vs. phonological syllables in handwriting production. *Cognition, 110,* 440-444.

Kandel, S., Soler, O., Valdois, S., & Gros, C. (2006). Graphemes as motor units in the acquisition of writing skills. *Reading & Writing, 19,* 313-337.

Kliegl, R., Masson, M. E., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18*(5), 655-681.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2016). *lmerTest: Tests in linear mixed effects models.* R package version 2.0-30. http://CRAN.R-project.org/package=lmerTest.

Lambert, E., Kandel, S., Fayol, M., & Espéret, E. (2008). The effect of the number of syllables when writing poly-syllabic words. *Reading & Writing: An Interdisciplinary Journal, 21,* 859-883.

Law, S. P. (1994). The structure of orthographic representations of Chinese characters: From the perspective of the cognitive neuropsychological approach. *Bulletin of Institute of History and Philology, 65,* 81–130.

Law, S. P. (2004). Writing errors of a Cantonese dysgraphic patient and their theoretical implications. *Neurocase, 10,* 132–140.

Law, S. P., & Caramazza, A. (1995). Cognitive processes in writing Chinese characters: Basic issues and some preliminary data. In B. de Gelder & J. Morais (Eds.), *Speech and reading: A comparative approach* (pp. 143–190). Hove, UK: Psychology Press.

Law, S. P., & Leung, M. T. (2000). Structural representations of characters in Chinese writing: Evidence from a case of acquired dysgraphia. *Psychologia, 43,* 67–83.

Law, S. P., Yeung, O., Wong, W., & Chiu, K. M. Y. (2005). Processing of semantic radicals in writing Chinese characters: Data from a Chinese dysgraphic patient. *Cognitive Neuropsychology, 22,* 885–903.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Lui, H. M., Leung, M. T., Law, S. P. & Fung, R. S. Y. (2010). A database for investigating the logographeme as a basic unit of writing Chinese. *International Journal of Speech-Language Pathology*, 12, 8-18.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29,* 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69–89.

O'Seaghdha, P.G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282–302.

O'Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: Attention constrains form preparation in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 797–810.

Orliaguet, J. P., & Boë, L. J. (1993). The role of linguistics in the speed of handwriting movements: Effects of spelling uncertainty. *Acta Psychologica, 82,*103–113.

Qu, Q., & Damian, M. F. (2015). Cascadedness in Chinese written word production. *Frontiers in Psychology, 6,* 1271.

Qu, Q. Q., Damian, M. F., & Li, X. (2016). Phonology contributes to writing: Evidence from a masked priming task. *Language, Cognition, and Neuroscience, 31*(2), 251-264.

Qu, Q. Q., Damian, M. F., Zhang, Q. F., & Zhu, X. B. (2011). Phonology contributes to writing: Evidence from written word production in a nonalphabetic script. *Psychological Science, 22,* 1107–1112.

R Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes: Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1130–1152.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35,* 854–876.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64,* 249–284.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb- particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 904–921.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14,* 173-200.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 13*(1), 33–7.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 922–939.

Shen, X. R., Damian, M. F., & Stadthagen-Gonzalez, H. (2013). Abstract graphemic representations support preparation of handwritten responses. *Journal of Memory and Language, 68*(2), 69-84.

State Language Commission. (1998). *The Chinese character component standard of GB13000.1 character set for information processing.* Beijing, China: Language & Culture Press.

Su, P. C. (1994). Xiandai Hanzixue gangyao [An introduction to the study of contemporary Chinese characters]. Beijing, China: Beijing University Press.

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language, 40,* 498–519.

Van Casteren, M. & Davis, M. H. (2006). Mix: A program for pseudorandomization. *Behaviour Research Methods, 38,* 584-589.

Zhang, P. (1984). Hanzi bujian fenxi de yuanjiu he lilun [The method and theory of analyzing logographemes in Chinese characters]. *Yuwenyanjiu, 1,* 37–43.

Zhang, Q., & Damian, M. F. (2012). Effects of orthography on speech production in Chinese. *Journal of Psycholinguistic Research, 41*, 267-283.

Zhou, X., & Marslen-Wilson, W. (1999). The nature of sublexical processing in reading Chinese characters*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 819–837.

Zhou, Y. G. (1978). 现代汉字中声旁的表音功能问题 [To what degree are the “phonetics” of present-day Chinese characters still phonetic?]. *Zhongguo Yuwen, 146,* 172–177.

Zhu, X. (1988). 现代汉字声旁表音功能的动态 [Analysis of cuing function of phonetic components in modern Chinese]. In X. Yuan (Ed.), *Proceedings of the symposium on the Chinese language and characters* (pp. 260–288). Beijing, China: Guang Ming Daily Press.

Table 1

Results of Experiment 1-4. Mean response latencies (RT, in milliseconds; standard deviations in parentheses) and mean error proportions (PE, in percent; standard deviations in parentheses), varied by context (heterogeneous, homogeneous, inconsistent).

	RT	Effect	PE	Effect
Experiment 1				
Heterogeneous	1026 (233)		1.6 (12)	
Homogeneous	957 (240)	69	3.1 (17)	-1.5
Experiment 2				
Heterogeneous	789 (239)		3.4 (18)	
Homogeneous	768 (265)	21	4.9 (22)	-1.7
Experiment 3				
Heterogeneous	779 (209)		2.2 (15)	
Homogeneous	785 (253)	-6	2.2 (15)	0
Inconsistent	775 (251)	4	2.4 (15)	-0.1
Experiment 4				
Heterogeneous	921 (233)		5.4 (23)	
Homogeneous	900 (242)	21	5.3 (22)	0.1

Appendix A

Materials used in Experiment 1

Context: Homogeneous

Set 1: 种植-粮食， 烘烤-糕点， 人才-精英， 优雅-粗鲁

Set 2: 脚趾-靴子， 过年-鞭炮， 敲诈-勒索， 射击-靶子

Set 3: 状元-科举， 动荡-稳定， 房价-租金， 可爱-稚气

Set 4: 拖延-耽误， 倾诉-聆听， 爽快-耿直， 工作-聘用

Context: Heterogeneous

Set 5: 种植-粮食， 脚趾-靴子， 状元-科举， 拖延-耽误

Set 6: 烘烤-糕点， 过年-鞭炮， 动荡-稳定， 倾诉-聆听

Set 7: 人才-精英， 敲诈-勒索， 房价-租金， 爽快-耿直

Set 8: 优雅-粗鲁， 射击-靶子， 可爱-稚气， 工作-聘用

Appendix B

Materials used in Experiment 2

Context: Homogeneous

Set 1: 豪宅-别墅， 军人-勋章，讽刺-鄙视， 长孙-嗣子

Set 2: 增添-剔除， 路灯-影子，放假-歇息， 答案-题目

Set 3: 亲人-故乡， 日本-韩国，打针-献血， 奋起-勃发

Set 4: 发明-创造， 街坊-邻居，西装-领带， 天空-鸽子

Context: Heterogeneous

Set 5: 豪宅-别墅， 增添-剔除， 亲人-故乡， 发明-创造

Set 6: 军人-勋章， 路灯-影子， 日本-韩国， 街坊-邻居

Set 7: 讽刺-鄙视， 放假-歇息， 打针-献血， 西装-领带

Set 8: 长孙-嗣子， 答案-题目， 奋起-勃发， 天空-鸽子

Appendix C

Materials used in Experiment 3

Context: Homogeneous

Set 1: 豪宅-别墅，军人-勋章，讽刺-鄙视，长孙-嗣子

Set 2: 增添-剔除，路灯-影子，放假-歇息，答案-题目

Set 3: 亲人-故乡，日本-韩国，打针-献血，奋起-勃发

Set 4: 算命-卦卜，绑架-劫持，地主-封建，宽恕-赦免

Context: Inconsistent

Set 5: 豪宅-别墅，军人-勋章，放假-歇息，答案-题目

Set 6: 增添-剔除，路灯-影子，讽刺-鄙视，长孙-嗣子

Set 7: 亲人-故乡，日本-韩国，地主-封建，宽恕-赦免

Set 8: 算命-卦卜，绑架-劫持，打针-献血，奋起-勃发

Context: Heterogeneous

Set 9: 豪宅-别墅，增添-剔除，亲人-故乡， 算命-卦卜

Set 10: 军人-勋章，路灯-影子, 日本-韩国，绑架-劫持

Set 11: 讽刺-鄙视，放假-歇息, 打针-献血，地主-封建

Set 12: 长孙-嗣子, 答案-题目, 奋起-勃发, 宽恕-赦免

Appendix D

Materials used in Experiment 4

Context: Homogeneous

Set 1: 离别, 功勋, 卑鄙, 子嗣

Set 2: 挑剔, 电影, 间歇, 标题

Set 3: 事故, 日韩, 奉献, 蓬勃

Set 4: 八卦, 打劫, 信封, 宽赦

Context: Heterogeneous

Set 1: 离别, 挑剔, 事故, 八卦

Set 2: 功勋, 电影, 日韩, 打劫

Set 3: 卑鄙, 间歇, 奉献, 信封

Set 4: 子嗣, 标题, 蓬勃, 宽赦

Figure Caption

Figure 1. Experiment 2. Top panel: response latencies for individual participants (1-24) dependent on Context. Dashed lines represent the model which assigns the same slope to all participants; solid lines represent the model which allows slopes to vary among participants. Bottom panel: response latencies for individual items (1-16). Dashed lines represent the model which assigns the same slope to all items; solid lines represent the model which allows slopes to vary among items.

Figure 2: Experiment 2. Left panel displays intercept adjustment for each participant based on overall speed, with the fastest participant at the bottom and the slowest on top. Right panel shows the corresponding Context effect as reflected in by-participant slope adjustments.

Figure 3. Experiment 4. Top panel: response latencies for individual participants (1-24) dependent on Context. Dashed lines represent the model which assigns the same slope to all participants; solid lines represent the model which allows slopes to vary among participants. Bottom panel: response latencies for individual items (1-16). Dashed lines represent the model which assigns the same slope to all items; solid lines represent the model which allows slopes to vary among items.

Figure 4: Experiment 4. Left panel displays intercept adjustment for each participant based on overall speed, with the fastest participant at the bottom and the slowest on top. Right panel shows the corresponding Context effect as reflected in by-participant slope adjustments.

Figure 1

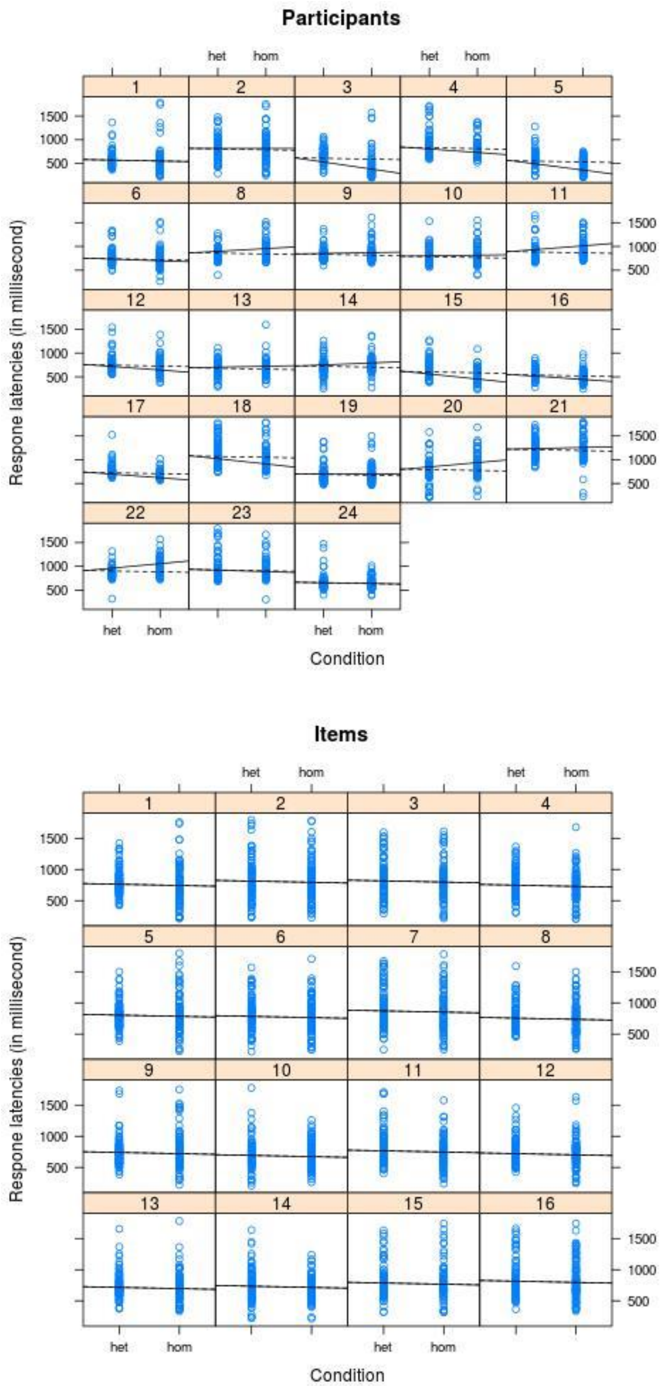
Participants
het
hom
Respone latencies (in millisecond)
Condition
Items
het
hom
Respone latencies (in millisecond)
Condition

Figure 2

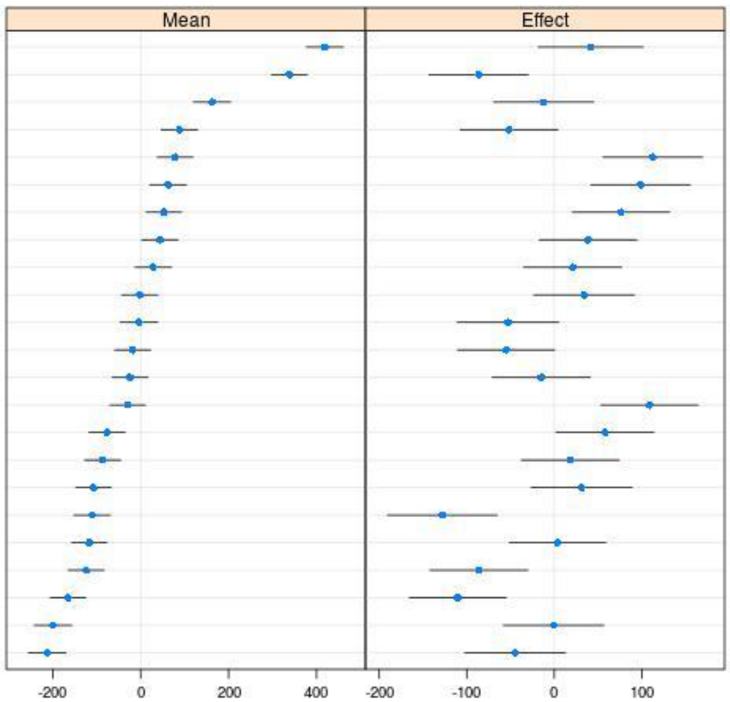

Figure 3

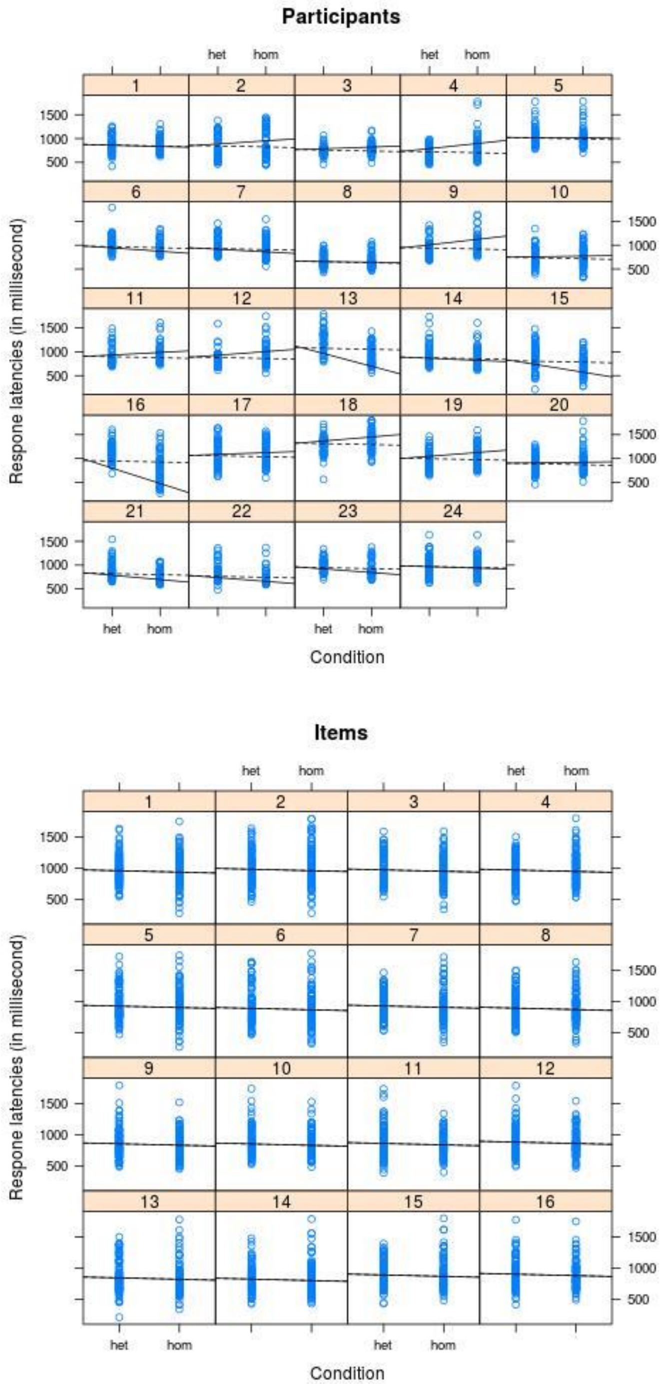
Participants
het
hom
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
1500
1000
500
Respone latencies (in millisecond)
Condition
Items
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
1500
1000
500
Respone latencies (in millisecond)
Condition

Figure 4